Available online at: http://e-jurnal.lppmunsera.org/index.php/JSMI

Jurnal Sistem dan Manajemen Industri
(Journal System and Management of Industries)

ISSN (Print) 2580-2887 ISSN (Online) 2580-2895

Early fault detection system for sugar mill machines through various machine learning approach

Thabed Tholib Baladraf[1*], Taufik Djatna[1], Agriananta Fahmi Hidayat[2], Akhmad Fatikhudin[1], Helynda Mulya Arga Retha[3], Zulfikar Dabby Anwar[4]

[1]Department of Agro-industrial Technology, IPB University, Jl. Raya Dramaga, Bogor, West Java 16680, Indonesia
[2]Department of Agriculture Engineering, Universitas Mataram, Jl. Majapahit No. 62, Mataram, West Nusa Tenggara 83115, Indonesia
[3]Department of Mathematics, Ludwig Maximilians Universität (LMU) München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany
[4]Faculty of Bioscience Engineering, Ghent University, Sint-Pietersnieuwstraat 33, 9000 Gent, Belgium

ARTICLE INFORMATION

Article history:

Received: May 19, 2025
Revised: September 26, 2025
Accepted: November 25, 2025

Keywords:

Early detection
Fault anomaly
Machine learning
Milling machine
Sugar agroindustry

*Corresponding Author

Thabed Tholib Baladraf
E-mail: thabedtholib@apps.ipb.ac.id

ABSTRACT

The milling machine is a crucial aspect of the sugarcane agroindustry production system; a disturbed milling machine will cause a decrease in production efficiency, sap quality degradation, and excessive energy consumption. An early fault anomaly detection system through machine learning is a solution to overcome the problems in sugarcane milling machines. The purpose of this research is to propose a system architecture design for early fault anomaly detection in sugarcane agroindustry milling machines and to evaluate the performance of various machine learning models on historical sensor data, identifying the most promising approach. This study proposes a novel anomaly detection framework for sugarcane milling machines to advance smart monitoring in agro-industrial systems. Using an empirical dataset of 7,673 sensor instances (temperature, vibration, pressure, and humidity), and applying several machine learning algorithms (logistic regression, decision tree, and random forest), the framework integrates multi-sensor data to improve fault prediction and reduce downtime. The results showed that the random forest had the best accuracy, at 98.13%, followed by the decision tree, at 97.87%, and logistic regression, at 89.70%. Feature contribution analysis reveals that the vibration signal is the most dominant contributing factor among other features. The results show that machine learning is a potential approach for predicting faults in sugarcane milling machines, which can help the sugarcane agriculture industry make informed decisions in the event of disturbances in these machines.

1. INTRODUCTION

Sugar is one of the compounds that has become a necessity for every human being. In its implementation, sugar is often used as a nutritional enhancer or as an additive to a food product [1]. In the process, sugar can be obtained from various plants that have a high carbohydrate content, especially those rich in sucrose. Some of the plants that can be used include juice, sugar palm, sugar beet, sago, coconut, and sugar cane [2]. Referring to the various sources of sugar raw materials, sugarcane is the most widely used crop in the process. According to Babu *et al*. [3], sugarcane contributes to 70% of the total sugar production worldwide, so it plays a very important role. Sugar produced from sugar cane plants, obtained through a series of processes, is considered to have quality and resources that are more abundant than those of other plants. Even in 2029, it is predicted to increase by 96% [4]. According to data from the United States Department of Agriculture, Brazil accounts for a 24% share of the global market, followed by India with 15%, the European Union with 9%, and China with 6% [5]. Apart from being used as a source of sugar, sugarcane also has potential for various other products, such as biomass, activated carbon, and particleboard [6]. This indicates that the sugarcane agroindustry has an increasing demand over time, necessitating more attention to operations.

Disrupted sugarcane agroindustry operations will certainly have an adverse impact on fulfilling sugar demand; one of the machines that has become a critical point is the sugarcane milling machine. According to Qiu *et al*. [7], the milling process is a crucial stage in obtaining juice from sugarcane. If the milling process can be optimized, it will lead to an efficient extraction process. Li *et al*. [8] stated that if the milling parameters can be optimized, it will provide good juice quality, optimal yield, and minimize energy costs. A disrupted sugarcane milling machine will halt the process or experience a slowdown, reducing the amount of sugarcane milled, and consequently, the sugar production target will not be achieved. Additionally, a malfunctioning milling machine can lead to a decrease in the quality of the juice. Sugarcane that is too late to undergo milling treatment will experience quality degradation in the form of a decrease in sugar content due to natural fermentation [9]. Delays in the milling handling process can also increase sugarcane impurities in the refining process, affecting the final product [10]. Negative impacts also arise in terms of the energy required, as the machine will work heavier than normal, resulting in greater energy production. There needs to be an early detection system for anomaly data through specific parameters on sugar cane milling machines, so that it can help the sugar agroindustry make informed decisions. Early detection of anomalies involves differentiating between abnormal patterns and normal patterns in data [11].

Several previous studies have been conducted to optimize the sugarcane milling process, aiming to achieve high-quality, high-yield, and low-energy juice. Meng *et al*. [12] used a kernel extreme learning machine to predict juice gravity purity and juice color value according to the criteria. Duan et al. designed a two-step method to determine the factors that have a significant influence on the milling process in sugarcane [13]. Nayak *et al*. [14] employed a machine fault simulator approach for rolling elements, including bearings, gears, belts, pulleys, and motor bearings, to enable the research to produce vibration pattern learning from the most common machine faults in a controlled manner without compromising production quality/profit.

However, there are still no reports on early fault anomaly detection systems in sugarcane milling machines. Based on this, an early fault anomaly detection system for sugarcane milling machines has been designed using important parameters such as temperature, pressure, vibration, and humidity, all of which are monitored through sensors. The data collected will be processed using machine learning with a supervised learning type, consisting of several variations of models, including logistic regression, decision trees, and random forests.

Logistic regression was chosen as one of the methodological approaches because it is simple and easy to use, especially for binary classification [15]. Logistic regression has efficient computation and can handle large datasets [16]. Moharam *et al*. [17] conducted research using several machine learning approaches, one of which was logistic regression for detecting anomalies in radio connection environments and obtained good accuracy results of 0.93.

Decision trees were also chosen as the approach used in this study because they provide a hierarchical decision tree approach to separate data based on feature values [18]. The ease of interpretation and robust performance are the advantages of decision trees, as the classification results are based on feature values and can control nonlinear relationships between the variables involved [19]. The use of decision trees in anomaly detection has been applied to other objects, such as spur gears and predicting machine downtime [20], [21].

The last model used was random forest, which is an ensemble approach that combines several decision trees to improve prediction accuracy. This enables random forest to handle data noise and overfitting [22], [23]. Kopp *et al*. [24] tested 34 anomaly datasets across various cases, achieving an impressive accuracy range. Additionally, Gao *et al*. [23] applied machine learning approaches, including random forest, in the iron industry for classifying damaged iron products.

The purpose of this research is to design an early fault anomaly detection system on sugarcane agroindustry milling machines and to find out the best machine learning approach for fault anomaly detection on sugarcane agroindustry milling machines. The main contributions of this research are described as follows:

1) This research introduces a novel anomaly detection framework for sugarcane milling machines, enhancing the current state of smart monitoring systems in the agro-industrial sector.
2) Offers a structured and detailed analysis of system requirements for anomaly detection in sugarcane agro-industrial machinery, thereby contributing methodologically to the software and systems engineering in agroindustry.
3) Through empirical experimentation, this study provides a comprehensive comparative assessment of machine learning approaches, thereby enriching the understanding of algorithmic suitability and performance in agro-industrial anomaly detection scenarios.
4) The proposed system architecture and findings serve as a strategic reference model for stakeholders in the sugarcane agroindustry to implement data-driven modernization and predictive maintenance practices, supporting the broader movement toward

http://dx.doi.org/10.30656/jsmi.v9i2.10530

Industry 4.0 in agricultural processing sectors.

This paper is organized in a coherent manner, where Section 2 will review similar previous studies related to anomaly detection in industrial processes. In Section 3, the proposed methodology is discussed, starting from requirement engineering, anomaly detection testing procedures, and anomaly detection evaluation. In Section 4, a discussion of the findings and their synthesis with previous research and theory is presented. In section 5, the conclusion of this research is presented in line with the research objectives.

2. RELATED WORK

Research related to anomaly detection in industrial machinery has been conducted extensively and offers various benefits. Anomaly detection research is considered capable of helping an industry recognize abnormal patterns in a working machine, enabling it to make informed decisions quickly and reduce downtime, thereby maximizing production. In anomaly detection, the indicators used are also an important consideration, as they must accurately represent the machine being detected.

Some previous research has been applied to various cases and industries using different approaches. There is still no anomaly detection in agricultural industrial machinery, specifically in the sugar cane agroindustry, so this research offers a novelty. Table 1 compares previous research in terms of industry, prediction approach used, and variables employed.

The majority of prior studies on anomaly detection have been conducted in manufacturing contexts, such as steel production, rotating machinery, bearings, gearboxes, and additive manufacturing. The majority of these studies focus on vibration, acoustic, or image-based signals, and while they achieve high accuracy with advanced models, their application remains concentrated on conventional industrial machines. Notably, none of the reviewed works address anomaly detection in the agro-industrial sector, particularly sugarcane milling machines, which present unique challenges due to their continuous operation, exposure to environmental variability, and complex multi-sensor dynamics. Moreover, prior studies often rely on laboratory or benchmark datasets rather than real-world industrial logs, limiting their practical applicability. To bridge this gap, our study introduces an anomaly detection framework specifically tailored for sugarcane milling machines, leveraging sensor data and grounding the labeling process in industry maintenance logs.

3. RESEARCH METHODS

This research will utilize three machine learning models with supervised learning types, namely logistic regression, decision trees, and random forests. In terms of how it works, logistic regression is a statistical technique used for binary classification problems, where the outcome variable is categorical (yes/no, 0/1). This model predicts the probability of an event occurring based on one or more predictor variables. The output is restricted to values between 0 and 1, representing probabilities. The logistic regression model employs the maximum likelihood method for parameter estimation, which involves an iterative process to determine the most suitable model [25], [26].

Meanwhile, a decision tree is a hierarchical model consisting of nodes that represent decisions or attribute tests, branches that represent the results of those tests, and leaf nodes that represent the final decision or classification. These trees are constructed using algorithms such as ID3, C4.5, and CART, which select the best attributes to split the data at each node based on criteria such as information gain or Gini impurity [27], [28]. On the other hand, random forest is an ensemble learning system that builds multiple decision trees and combines their predictions to improve accuracy and control overfitting. Each tree in the forest is trained on a random subset of the data and a random subset of the features, which introduces diversity among the trees. The final prediction is made by averaging the predictions using majority voting for classification [29], [30].

The three models used will be trained and tested using a dataset of 7,673 data consisting of temperature, pressure, humidity, and vibration measurements, with the data labeling process conducted based on the maintenance logs provided by the sugar agro-industrial. Anomalies in sugar cane milling machines that will be detected include various suspected damages, such as electrical damage and thermal damage. The sensors used are considered capable of representing the machine's behavior completely.

The dataset is obtained from sensors installed on a sugar cane milling machine with 5 rollers and a 7,000 ton per day (TCD) capacity. A total of 70% of the collected data will be used for training, and 30% will be used for testing, according to the three models employed: logistic regression, decision tree, and random forest. The data that has been tested will then be evaluated using several measurement metrics, especially from the accuracy and goodness model. In real-world implementation, the system is real-time because it functions in early detection and is implemented on a real scale.

3.1. Vibration parameter

The vibration aspect in an industry is one of the important indicators in predictive maintenance practices. In this study, the sensor used was a YDS106 type. Vibration monitoring has proven to be an effective method for finding damage to machine components [31], [32], [33]. In the diagnosis of machine operation using vibration parameters, ISO 22096:2007

Table 1. Comparison with previous research

No	Authors	Scope	Contribution	Performance	Variable
1.	Alcazar *et al.* [34]	Construction tools industry	The research contributes to providing a framework for security solutions in industrial using supervised learning for real time anomaly detection in the construction industry.	SVM (92.3%), RF (91.5%), CNN (94.2%)	Temperature, flow, speed, pressure.
2.	Gamal *et al.* [35]	Steel plate product industry	The research contributes in providing an anomaly detection solution by supervised learning in steel plates.	DT (91.1%), KNN (82.9%), RF (92.9%), SVM (86%), LR (88.3%), MLP (73.9%)	Pixel areas, perimeter, log areas, length of conveyor, type steel, thickness, luminosity.
3.	Huang *et al.* [36]	Rotating machinery chemical industry	The findings reveal intelligent fusion that combine the favorable characters of different to drive the development of fault diagnosis prediction	BNN (91.6%), ENN (27.2%), RBFN (83.6%), PNN (66%), WNN (84.8%)	Accelerometers
4.	Das & Das [37]	Rotating machinery	This paper introduces genetic algorithm to optimized boosted trees for fault identification in rotating machinery.	GADA (99.8%)	Tachometer, sounds, and accelerometers
5.	Wang *et al.* [38]	Bearing machinery chemical industry	This paper proposes a novel hybrid approach of a wavelet packet and random forests classifier for the fault diagnosis in rolling bearings.	WPDRF (88.2%)	Vibration
6.	Li *et al.* [39]	Gearbox industry	This work addresses the use of a deep random forest fusion to fault diagnosis for gearboxes by using acoustic and accelerometer.	DRF (97.7%)	Acoustic & vibration
7.	Chow *et al.* [40]	Concrete material industry	This implementing deep learning for anomaly detection of defects on concrete structures.	DNN (65.87%)	Fusion
8.	Scime & Beuth [41]	Laser machinery	This paper proposes autonomous detection of many anomalies of defects in laser powder bed fusion	CNN (85%)	Fusion
9.	Cooper *et al.* [42]	Milling machinery metal	This paper demonstrates the detection of anomalies in the time-frequency domain of the tool's acoustic spectrum during cutting operations.	GAN (90.56%)	Acoustic
10.	Mattera & Nele [43]	Wire arc additive	This paper compares unsupervised, supervised, and semi-supervised approaches with small datasets in wire arc additive manufacturing.	LR (95.3%), RF (95.7%), DT (93.7%), CNN (96.8%)	Voltage signals
11	Proposed Research	Milling machinery sugarcane agroindustry	This study develops and validates a machine-learning-based anomaly detection framework for sugarcane milling machines, integrating requirement analysis, model comparison, and system architecture.	LR (89.70%), DT (97.87%), RF (98.13%)	Vibration, temperature, humidity, pressure

Note: SVM = support vector machine, RF = random forest, CNN = convolutional neural networks, DT = decision tree, KNN = k-nearest neighbor, NB = naïve bayes, LR =logistic regression, MLP =multi-layer perception, BNN = backpropagation neural network, ENN = elman neural network, RBFN = radial basis function neural network, PNN = probabilistic neural network, WNN = wavelet neural network, GAADA = genetic algorithm and adaboost, WPDRF = wavelet packet denoising random forest, DRF = deep random forest, DNN = deep neural network, GAN = generative advers network

 http://dx.doi.org/10.30656/jsmi.v9i2.10530

is the reference. Vibration is the oscillatory motion of equipment around its equilibrium position. In the context of vibration, any change in the amplitude or frequency of the signal indicates that the machine performance is impaired [44], [45]. Vibration analysis can be an effective tool for diagnosing looseness, eccentricity, imbalance, blade defects, misalignment, defective bearings, damaged gears, and cracked or bent shafts [46], [47]. In practice, predictive maintenance has various advantages compared to other parameters, including high accuracy, sensitivity to various types of defects, and being a non-invasive or non-destructive method [48], [49].

3.2. Temperature parameter

Temperature is one of the important parameters in predictive maintenance prediction and is usually described through fluctuations in the running machine. In this study, the sensor used is a heat-resistant RTD PT100. Temperature can provide an overview of potential problems that occur in a machine through several assumptions, such as overheating or inefficiency [50]. In a milling machine, temperature can describe the performance of the machine, whether it is in accordance with its capacity and whether it is receiving a sufficient electricity supply. Stable temperature, in accordance with the criteria, is closely related to the performance of an industrial machine. Monitoring temperature fluctuations is crucial to ensure operational efficiency, safety, and equipment longevity [51]. Temperature anomalies, if not detected, can cause equipment damage, production disruptions, and safety hazards [52].

3.3. Pressure parameter

The pressure on the milling machine is used for measurement in relation to the pump on the milling machine after the Nira is squeezed. In this study, the sensor used is a P20T. The juice that has been produced will be pumped to the next machine for further processing. Pump pressure that does not meet the criteria will reduce the yield channeled to the next process [53]. This affects the sugar yield in the final product. A too-high-pressure sensor is indicative of an overload phenomenon, while a low-pressure sensor is indicative of a leak in the component [54], [55]. Pressure sensors also do not directly cause total damage to the milling machine; however, they can serve as an early indicator or early detection of major damage [56].

3.4. Humidity parameter

Excessive humidity in the sugarcane milling machine environment can be a significant factor contributing to various forms of systemic damage. In this study, the sensor used is an SHT85. High humidity accelerates the oxidation process of metals, triggering corrosion in vital components such as bearings, shafts, gears, and electrical connectors. This ultimately reduces efficiency, accelerates mechanical wear, and increases the risk of total machine failure [57]. In addition, modern sugarcane milling machines equipped with electronic systems and digital sensors are highly susceptible to condensation due to high humidity, which can cause short circuits, insulation disorders, and even total sensor failure [58]. Excess humidity also encourages the growth of microorganisms that contaminate processed products, thereby reducing product quality. Therefore, humidity sensors play a crucial role as anomaly indicators in predictive maintenance systems. Changes in humidity trends from historical values can indicate damage such as cooling system leaks, damaged seals, or abnormal changes in the operating environment [59].

3.5. Evaluation of prediction performance

To measure prediction performance, several evaluation metrics will be used, including the confusion matrix, the area under the curve-receiving operating characteristic (AUC-ROC), and the mean squared error (MSE). In the confusion matrix evaluation, performance will be measured using accuracy, precision, sensitivity, and F1-score. The accuracy, precision, sensitivity, and F1-score values are obtained from a 2 x 2 table that compares correct answers and incorrect answers. In the confusion matrix, there are true positive, true negative, false positive, and false negative. True positive and true negative are measurement results that indicate a direct correlation between the actual value and the predicted value, while false positive and false negative are measurement results that show a non-linear relationship between the actual value and the predicted value.

Evaluating prediction performance, there are also AUC-ROC and MSE, which are used to determine the goodness of the model in predicting classification. In measuring AUC-ROC, there is a standard: if the value is close to 1, then the model can be said to be improving [54]. Conversely, in MSE, the smaller the number produced, the better [60]. The details of the measurement formulas, including accuracy, precision, sensitivity, F1-score, AUC-ROC, and MSE, are presented as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Sensitivity = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$F1\ score = \frac{2(sensitivity\ x\ precision)}{(sensitivity + precision)} \tag{4}$$

$$Mean\ Squared\ Error = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

$$AUC - ROC = \int_0^1 TPR\left(FPR^{-1}(t)\right)dt \quad (6)$$

The prediction performance evaluation system is a crucial component in assessing the accuracy of a model, particularly in classification tasks. Several commonly used evaluation metrics include accuracy, precision, and sensitivity. In this context, several standard notations are applied: TP (true positive) refers to the number of positive cases correctly classified by the model; FP (false positive) refers to negative cases incorrectly classified as positive; TN (true negative) represents the number of negative cases correctly identified; and FN (false negative) refers to positive cases that are incorrectly classified as negative. Meanwhile, the mean squared error (MSE) formula involves the notation n, which indicates the total number of observations in the dataset. The notation y_i represents the actual value of the *i-th* observation, while $\hat{y}_i$ denotes the predicted value for that same observation. For more advanced classification model evaluation, such as the area under the curve - receiver operating characteristic (AUC-ROC), the notations TPR (true positive rate), equivalent to sensitivity, and FPR (false positive rate), which refers to the proportion of actual negative cases incorrectly classified as positive, are used. Lastly, *dt* denotes the differentiation with respect to variable *t*, commonly used in integral calculations within ROC curve analysis.

3.6. Research stages

The research began with the collection of temperature, pressure, vibration, and humidity data obtained from sensors installed in sugarcane milling machines to monitor the performance of the sugarcane milling machine. The data collected from the sensors is raw data that is then checked for missing values, cleaned, and normalized.

The next step is to verify and clean the data to eliminate empty values, ensuring it is ready for analysis. At this stage, it involves checking the data for validity. The handling of missing data was performed using simple imputation, specifically filling in previous values (forward fill). Normalization was then applied using minimum-maximum normalization, ensuring all values fell within the range of 0-1 for subsequent data processing. Sensor observation data was collected 21 times a day over a period of 1 year, as the dataset used.

Data that is clean, normally distributed, free from multicollinearity issues, and not affected by autocorrelation phenomena will be processed further in the analysis stage using three selected machine learning models: logistic regression, decision trees, and random forest. Each of these models was chosen because it represents different approaches to classification, ranging from statistical probability-based methods to rule-based learning and ensemble techniques. After the models are trained and tested, the results of the analysis will be carefully compared using several evaluation metrics, particularly the confusion matrix and additional statistical measures, to ensure the reliability and validity of the findings. Once the comparative evaluation is complete, the results will undergo a deeper investigation through feature importance analysis, which aims to identify the most significant variables that influence the objectives of the study. This step provides practical insights for interpretation and decision-making. The complete stages of this research process are systematically illustrated in Fig. 1.

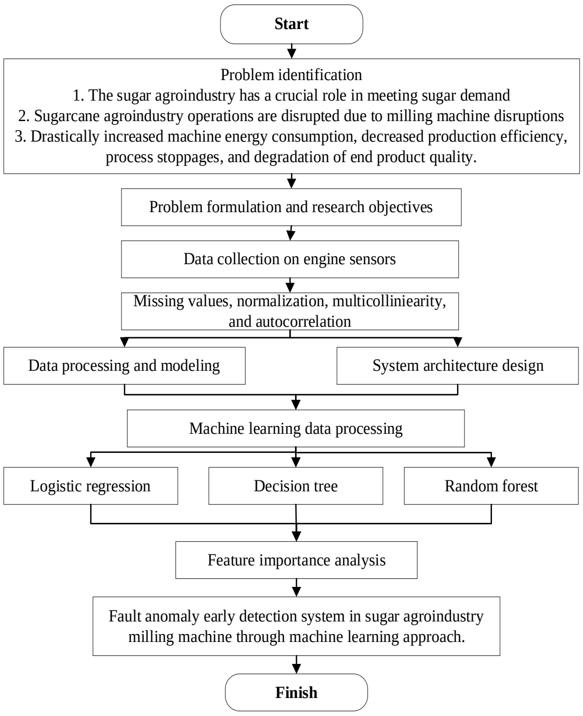

Fig 1. Research stages of sugarcane milling machine anomaly prediction through machine learning

4. RESULTS AND DISCUSSION

4.1. Early fault detection system development

The development of a system is fundamentally determined by the specific requirements of the industry in which it will be applied. In this research, the requirements are categorized into three primary aspects: technical, functional, and non-functional. These aspects are further visualized through block diagrams that illustrate abstract components and their interrelationships. The diagrams highlight essential system elements, including sensors, data processing units, feedback mechanisms, and system control. Focusing on the sugarcane milling machine as the primary object of study, this research emphasizes the need for accurate, real-time monitoring of critical subsystems. To achieve this, a set of sensors (temperature, humidity, vibration, and pressure) is deployed to collect relevant operational data from different stages of the milling process.

The sensor data is first transmitted to a centralized database to ensure secure and structured storage. Once

 http://dx.doi.org/10.30656/jsmi.v9i2.10530

stored, the data is processed through a data control system where advanced analysis techniques are employed to detect anomalies and predict potential faults. These insights are expected to help decision-makers maintain efficiency, minimize downtime, and prevent costly machine failures. In doing so, the system serves as an intelligent support tool for operational management in the agroindustry. A detailed overview of the system requirements designed for anomaly detection in the sugarcane milling process is provided in Table 2.

Table 2. System requirement

Requirements	Detail
Technical	1. Temperature, vibration, pressure, and humidity sensor (data act system) 2. Technician, connector data, and operator (data processing & control) 3. Data visualization and reporting system
Functional	1. The system can collect data through the sensors used in real time 2. The system can analyze anomaly detection data using machine learning 3. The system can provide warning notifications to stakeholders in real time 4. The system can display data in an interactive dashboard of machine monitoring
Non-functional	1. The system is responsive and has fast latency 2. The system has high accuracy in detection 3. The system must be scalable or adaptive 4. The system has data encryption to protect information. 5. The system has a user interface friendliness and fully documented.

The block diagram of the fault anomaly detection system in the sugarcane agro-industrial milling machine will be divided into three parts: data acquisition, data processing, and action. All these parts are interconnected and have their respective roles. Data acquisition is tasked with collecting data through sensors installed on the machine. Data processing and control will store and process data to predict fault anomalies, producing a report. Action is taken based on the decision-making process, informed by the prediction results obtained.

The proposed system approach is also expected to be scalable, allowing it to be applied to other sugar cane mills, as there are similarities in damage indicators between one sugar cane mill and another. The details of the fault anomaly detection object abstraction at milling machine are presented in Fig. 2.

4.2. Data processing and modelling

Data that has been collected and will be used in machine learning modeling needs prerequisite testing to validate the data used. The prerequisite tests used in this case are multicollinearity testing and auto-correlation testing. Multicollinearity testing is conducted to determine the correlation between independent variables, as measured by the variance inflation factor (VIF). Data is said to have high multi-collinearity if it produces VIF> 10, said to be moderate multicollinearity if VIF 5-10, and said to be low multicollinearity if VIF<5 [61]. In addition to multicollinearity, there is an autocorrelation test, which is a condition of checking the correlation between members of a series of observations sorted by time or space. In this study, the Durbin-Watson test will be used, with an indicator close to 2. If the Durbin-Watson (DW) value is closer to 2, it indicates the absence of autocorrelation [62].

The results of multicollinearity testing found that all data used in the test belonged to low multicollinearity. In more detail, temperature gets a VIF value of 1,006, pressure of 1,008, vibration of 1,010, and humidity of 1,000. Meanwhile, the autocorrelation prerequisite test shows that there is no autocorrelation because it gets a DW value of 1.969. Based on the prerequisite testing obtained, it is known that the data obtained is suitable for further use in machine learning testing. The data will then be prepared and divided for training and for testing, roughly 70% of the data (5,371 data) and 30% of the data (2,302 data).

4.3. Prediction through logistic regression, decision tree, and random forest

To evaluate the feasibility of using machine learning for anomaly detection within the proposed system context, several models were trained and tested using historical sensor data collected from the relevant machines. Based on the analysis that has been done, it is found that random forest has the best prediction ability compared to other models, namely logistic regression and decision tree. Random forest is a better model because this model can handle overfitting problems in the prediction process. Random forest can bagging or bootstrap aggregating is an ensemble learning technique that performs resampling in parallel and combines the results to produce a final prediction.

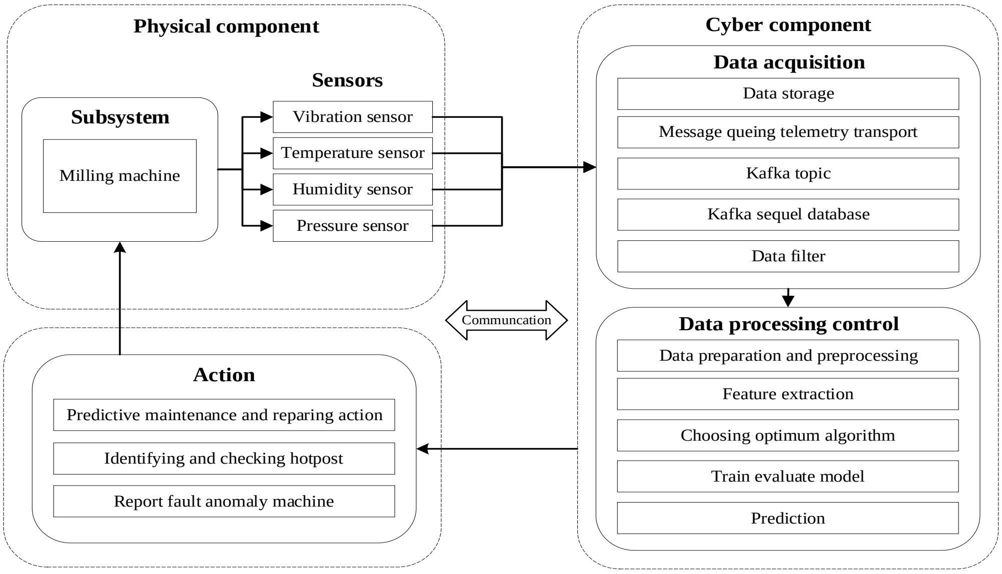

Fig 2. Proposed system architecture concept

Random forest has the ability to obtain a globally optimal solution, this makes random forest as an ensemble model superior to a single model [63]. On the other hand, random forests are also able to reduce bias and variance, this is because there is a merging of several trees so that it can produce low-bias and low-variance prediction results. These results are in line with research conducted by Gamal *et al*. [35], which shows that random forest is also the best approach for predicting damage to steel plates with an accuracy of 92.9%. Mattera & Nele [43] also showed similar results, making the random forest approach the best approach with an accuracy of 95.7% in predicting safety in the manufacturing industry, capable of detecting data manipulation, overload, and intrusion. Furthermore, Li *et al*. [39] developed a model using deep random forests to predict anomalies in gearboxes and achieved an excellent accuracy of 97.7%. On the other hand, Wang *et al*. [38] conducted development by combining wavelet packet and random forests for predicting anomalies in chemical industry gear bearings with an accuracy of 88.2%.

Decision tree also has a performance that is not much different from random forest, this is because the decision tree has a similar working principle to random forest. However, in creating a solution, the decision tree only considers local solutions without considering other branches of the tree. Research conducted by Mattera & Nele [43] in predicting the security of manufacturing industry data shows an accuracy that is not significantly different from that of random forest, which is 93.7%. Halabaku & Bytyci [64] stated that the decision tree becomes a relatively sensitive model, capturing noise as a meaningful pattern, thus reducing the robustness of the decision tree model. Additionally, the decision tree, as a single model, often experiences overfitting in the prediction process. As a result, the complete prediction and comparison results of the anomaly fault detection system on the sugarcane agroindustry milling machine are presented in Table 3 and Fig. 3. Logistic regression is a machine learning model that has lower accuracy compared to decision trees and random forests. Logistic regression is one of the models that has limitations in capturing nonlinear patterns between predictor variables (temperature, humidity, vibration, and pressure sensor data) and the log of the target variable (fault anomaly detection). This finding aligns with Borucka & Grzelak [65], which demonstrates that the use of logistic regression in evaluating the efficiency of production machines yields an accuracy of only 67.4% with an AUC value of 0.7156.

Table 3. Comparison of the performance of fault anomaly prediction systems in sugar agroindustry milling machines

Machine Learning	MSE Train	MSE Test	Accuracy Train	Accuracy Test	F1 Score	AUC Score	Precision Score	Recall Score
LR	0.098	0.102	90.13%	89.70%	84.84%	0.719	44.85%	50.00%
DT	0.016	0.021	98.34%	97.87%	97.79%	0.906	97.69%	90.60%
RF	0.000	0.0186	100%	98.13%	98.11%	0.973	95.98%	93.73%

Note: Machine learning approach (LR = Logistic Regression, DT = Decision Tree, RF = Random Forest)

 http://dx.doi.org/10.30656/jsmi.v9i2.10530

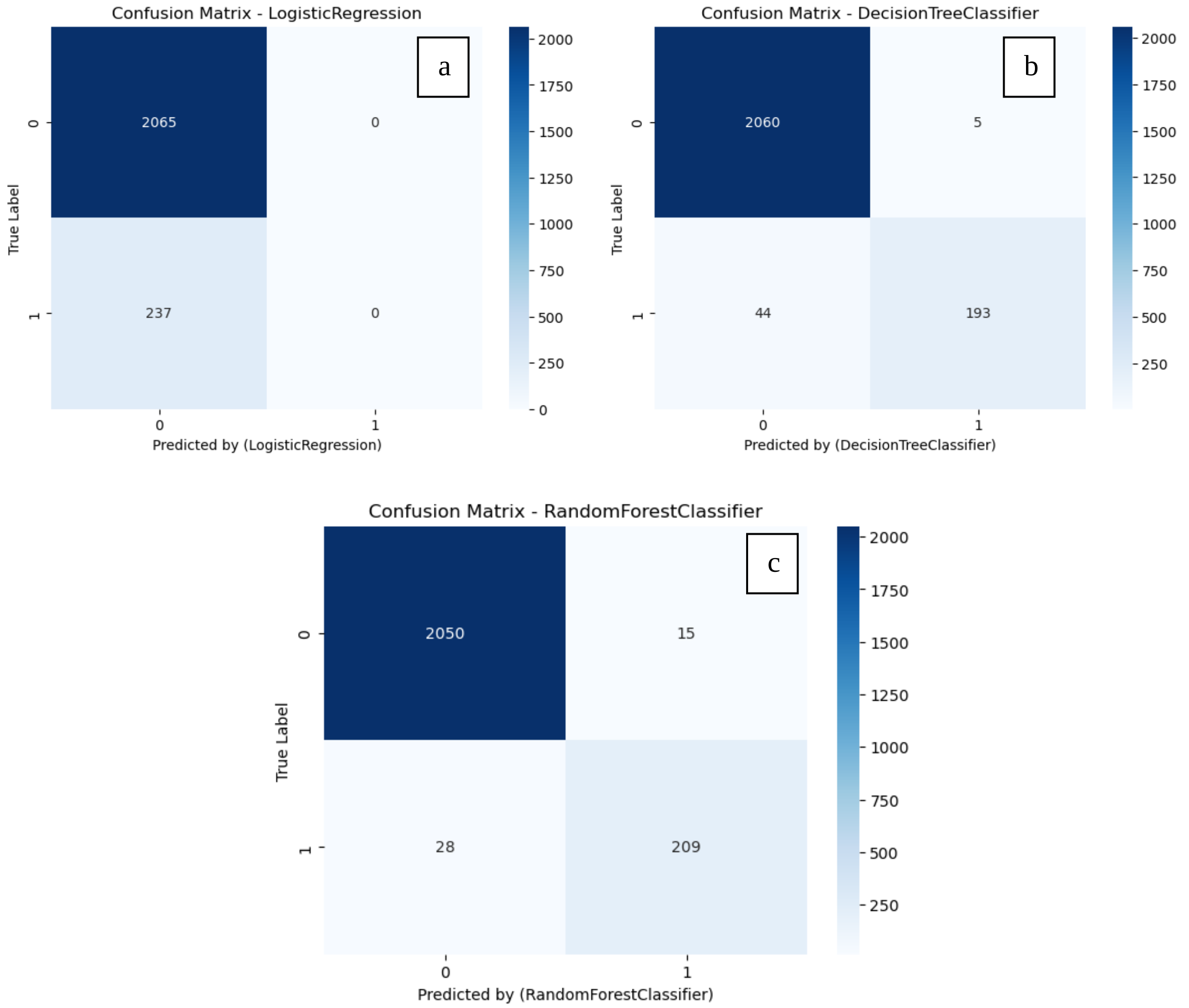

Fig 3. Confusion matrix (a) logistic regression, (b) decision tree, and (c) random forest

On the other hand, logistic regression is also difficult to identify complex interactions that occur between variables. The interaction between predictor variables in this case is classified as a complex interaction because it has four predictor variables. This is further exacerbated by the sensitivity of logistic regression to noise; the presence of noise in the dataset will reduce the performance of logistic regression. Research by Gamal *et al.* [35] shows that logistic regression is not superior to random forest and decision tree in predicting steel plate damage.

This research presents a new approach to digital transformation, particularly in the early detection of agroindustry machinery. Most early detection research on machinery is conducted in another industry with different conditions, such as domain-specific applications, source data, evaluation setups, and high complexity [66], [67], [68]. Research by *Hu et al.* [66] was conducted to detect damage to bearing components in the petrochemical industry using vibration signals as a single predictor variable. The anomaly detection system built has a relatively good accuracy of 95.58%. Wen et al. conducted fault detection on screw pumps in the petrochemical industry using several indicators as predictor variables, such as voltage, power, load, torque, rotational speed, and pressure. The results of the study have a fairly high accuracy, ranging from 69.5% to 75.6% [69].

In comparison with the manual inspection system implemented by sugar factories, there are various advantages, including the fact that the system is more proactive than manual inspections, which are reactive because decisions are only made after symptoms appear. Machine learning and sensor-based inspection systems are capable of providing real-time checks, enabling regular monitoring of machine conditions and eliminating human subjectivity, as they have learned from machine behaviour. The machine learning and sensor-based inspection system can provide an overview of complex symptom relationships that cannot be captured through manual inspection due to the interplay between multiple variables involved.

4.4. Feature importance

Feature importance is a crucial stage in identifying the most important or contributing features to a system's design. In this research, feature importance is performed at the end to determine how the system makes decisions, as well as to validate the results. The results of feature importance are presented in Table 4.

Table 4. Features an important analysis

Feature	LR	DT	RF
Vibration	2.170	0.505	0.330
Temperature	0.689	0.394	0.366
Pressure	0.594	0.074	0.185
Humidity	0.229	0.025	0.111

Note: Machine learning approach (LR = Logistic Regression, DT = Decision Tree, RF = Random Forest)

The vibration feature is the most dominant aspect, followed by temperature, pressure, and humidity. Vibration is an aspect that is synonymous with disruption in a machine; the higher the vibration value produced, the higher the potential for the machine to experience disruption. Vibration is also useful, especially in the early detection of machine failure. This is because vibration signals can provide signs before other symptoms appear, such as changes in temperature, pressure, humidity, and sound. Even specific vibration signals can provide diagnostic capabilities regarding the type of damage experienced by the machine [70].

Previous research from Tambake *et al*. [71] performed diagnosis on CNC machines using vibration features of three axes as predictor variables and obtained perfect prediction results. Yuan *et al*. [72] also utilised vibration signals from rotors, gearboxes, and rolling bearings to detect rotating machine failures, achieving results with good accuracy. This demonstrates that vibration signals are a crucial indicator in failure detection.

4.5. Proposed corrective actions

The proposed system emphasizes concrete corrective actions to mitigate failures once they are detected. This is essential because anomaly prediction alone, without actionable responses, would limit the system's utility for industrial operations. Each anomaly indicator corresponds to a specific operational issue and requires targeted intervention. For instance, abnormal vibration patterns may signal bearing wear, roller imbalance, or shaft misalignment; in such cases, corrective measures such as bearing replacement, rotor rebalancing, and alignment adjustments should be carried out immediately. Elevated temperature readings can indicate lubrication failure, inefficient cooling, or electrical overload, requiring operators to inspect the lubrication system, verify cooling mechanisms, and evaluate motor performance. Likewise, abnormal pressure values may suggest leakage, pump overload, or seal degradation, which can be remedied through pump inspection, seal replacement, and pipeline maintenance. High humidity levels, often associated with corrosion and sensor malfunction, can be mitigated by ensuring adequate ventilation, repairing faulty seals, and implementing dehumidification procedures [73]. These corrective actions directly connect to the anomaly detection outputs, ensuring that the insights generated by the model lead to tangible operational improvements.

To ensure practical implementation, the anomaly detection system should be integrated with a real-time alert mechanism that categorizes deviations into early warnings and critical alerts. This aligns with the predictive framework previously described, where models such as random forest not only achieve high classification accuracy but also provide probability-based outputs that can be mapped to risk thresholds. Early warnings enable operators to schedule maintenance interventions without disrupting production, while critical alerts trigger immediate inspections and corrective actions to prevent system breakdowns [74]. In this way, the detection system evolves from a passive classification tool into a proactive decision-support mechanism. Furthermore, by structuring corrective actions into condition-based maintenance (CBM) strategies, the system enables a shift from traditional scheduled maintenance toward predictive and adaptive maintenance routines. This integration bridges the gap between predictive analytics and operational practice, ensuring that anomaly detection translates into operational resilience.

By linking anomaly detection results with actionable protocols, the proposed framework ensures that failures are not only identified but also systematically addressed in alignment with the predictive system architecture described earlier. The corrective actions serve as the execution layer of the anomaly detection pipeline, reinforcing the value of predictive insights by preventing unplanned downtime, extending equipment life, and improving product quality. Moreover, documenting each corrective intervention in maintenance logs creates a feedback loop that strengthens the anomaly detection models [75]. Historical corrective action data can be reintegrated into the training process, refining model accuracy, improving feature importance interpretation, and supporting more efficient resource allocation. This closed-loop system positions the proposed framework as not only a predictive tool but also as a continuous improvement mechanism for smart agro-industrial operations.

4.6. Research implications

This study has considerable implications for the cane sugar agroindustry, producing operational

http://dx.doi.org/10.30656/jsmi.v9i2.10530

multiplier effects that extend beyond mere enhancements in monitoring. The system emphasises the strategic importance of leveraging historical and real-time data assets, a benefit that becomes increasingly pronounced for mills that already utilise digital record-keeping systems. The integration of these datasets enables advanced capabilities, including machine life prediction, workflow diagnostics, and enhanced process optimization. Its high-accuracy anomaly detection module serves as a practical decision-support tool that expedites maintenance responses, reduces diagnostic uncertainty, and shortens the duration of repair interventions [76].

The earlier and more reliable identification of faults enables mills to transition from rigidly scheduled maintenance to condition-based maintenance, thereby directly reducing downtime, minimizing operational waste, and stabilizing throughput, while enhancing juice extraction efficiency and quality. Beyond these operational improvements, the AI-based framework provides a tangible modernization pathway for an industry that has remained technologically stagnant for centuries. It strengthens mill competitiveness, facilitates scalable digital transformation, and offers a replicable model for data-driven innovation across agro-industrial processing sectors [77].

5. CONCLUSION

This research develops a damage anomaly detection system for sugarcane milling machines to enhance production efficiency, particularly during the sap extraction process, while also preventing production stoppages, preserving product quality, and reducing energy consumption. By rigorously evaluating several machine learning models using historical sensor data, the random forest model emerged as the most accurate, achieving 98.13% accuracy with an average deviation of 1.86%. The decision tree achieved an accuracy of 97.87%, whereas logistic regression showed the lowest performance at 89.70%, primarily due to its limitations in modeling non-linear relationships and variable interactions. This study addresses a significant research gap in the sugarcane agroindustry, as integrated machine learning-based anomaly detection systems have not yet been widely adopted. The research contributes to the field by both assessing the effectiveness of predictive models and proposing a system architecture suitable for implementation in sugar mill operations. The application of historical sensor data for fault detection represents a relatively novel approach in this industrial context. The findings align with the study's objectives by demonstrating high prediction accuracy and presenting a viable system design that offers tangible solutions to pressing issues related to efficiency, product quality, and energy usage.

Nevertheless, the approach has limitations, including testing conducted within a single mill environment, which affects the model's generalizability. The system has not yet been validated in real-time conditions, and its economic viability has not been analyzed. Future work will focus on integrating real-time monitoring with the Internet of Things, exploring advanced deep learning models such as long short-term memory and autoencoders, incorporating explainable artificial intelligence to support technician decision-making, and evaluating the system's economic and environmental impacts.

ACKNOWLEDGMENT

The authors gratefully acknowledge the PMDSU scholarship support from the Ministry of Education, Culture, Research, and Technology of Indonesia and IPB University. The authors also express sincere appreciation to Prof. Dr Ir. Marimin, M.Sc., for the valuable knowledge and insights he has shared, as well as for the opportunity to learn and grow within the Department of Industrial Technology at IPB University.

REFERENCES

[1] J. Walton, H. Bell, R. Re, and A. P. Nugent, "Current perspectives on global sugar consumption: definitions, recommendations, population intakes, challenges and future direction," *Nutr. Res. Rev.*, vol. 36, no. 1, pp. 1–22, Jun. 2023, doi: 10.1017/S095442242100024X.

[2] H. Erikliogu *et al.*, "Classification and quantification of sucrose from sugar beet and sugarcane using optical spectroscopy and chemometrics," *J. Food Sci.*, vol. 88, no. 8, pp. 3274–3286, Aug. 2023, doi: 10.1111/1750-3841.16674.

[3] K. S. Dinesh Babu, V. Janakiraman, H. Palaniswamy, L. Kasirajan, R. Gomathi, and T. R. Ramkumar, "A short review on sugarcane: its domestication, molecular manipulations and future perspectives," *Genet. Resour. Crop Evol.*, vol. 69, no. 8, pp. 2623–2643, Dec. 2022, doi: 10.1007/s10722-022-01430-6.

[4] D. Medina Hidalgo, A. Mallette, S. Nadir, and S. Kumar, "The future of the sugarcane industry in Fiji: climatic, non-climatic stressors, and opportunities for transformation," *Front. Sustain. Food Syst.*, vol. 8, no. March, pp. 1–13, Mar. 2024, doi: 10.3389/fsufs.2024.1358647.

[5] USDA, "Sugar : World Markets and Trade," 2025, *USDA, United States*. [Online]. Available: https://apps.fas.usda.gov/psdonline/circulars/sugar.pdf

[6] N. M. Salatein, R. A. Ibrahim, and I. S. Fahim, "Sustainable utilization of sugarcane bagasse for wood-based panels : A promising approach for waste management in Egypt Sustainable utilization of sugarcane bagasse for wood-based panels : A promising approach for waste management in Egypt," *J. Eng. Res.*, no. May, 2024, doi: 10.1016/j.jer.2024.05.013.

[7] M. Qiu, Y. Meng, J. Chen, Y. Chen, Z. Li, and J. Li, "Dual multi-objective optimisation of the cane milling process," *Comput. Ind. Eng.*, vol. 179, p. 109146, May 2023, doi: 10.1016/j.cie.2023.109146.

[8] Z. Li *et al.*, "Multi-Objective Optimization of Sugarcane Milling System Operations Based on a Deep Data-Driven Model," *Foods*, vol. 11, no. 23, p. 3845, Nov. 2022, doi: 10.3390/foods11233845.

[9] P. Dhansu *et al.*, "Different Treatments for Sugarcane Juice Preservation," *Foods*, vol. 12, no. 2, p. 311, Jan. 2023, doi: 10.3390/foods12020311.

[10] N. Ayele, T. Tana, P. D. R. van Heerden, K. W/Tsadik, and Y. Alemayehu, "Sugarcane Yield and Juice Quality as Affected by Harvest Age and Chemical Ripeners in Central Rift Valley of Ethiopia," *Agric. Res.*, vol. 12, no. 4, pp. 421–427, Dec. 2023, doi: 10.1007/s40003-023-00664-9.

[11] R. J. Nayak and J. P. Chaudhari, "Anomaly detection using deep learning based model with feature attention," *IAES Int. J. Artif. Intell.*, vol. 13, no. 1, p. 383, Mar. 2024, doi: 10.11591/ijai.v13.i1.pp383-390.

[12] Y. Meng, S. Yu, H. Wang, J. Qin, and Y. Xie, "Data-driven modeling based on kernel extreme learning machine for sugarcane juice clarification," *Food Sci. Nutr.*, vol. 7, no. 5, pp. 1606–1614, May 2019, doi: 10.1002/fsn3.985.

[13] Q. Duan, H. Mao, H. Mao, Z. Huang, B. Li, and X. Li, "Study on elastoplastic coupling mechanics model of the milled mixture of sugar cane," *Food Sci. Technol.*, vol. 39, no. suppl 1, pp. 270–277, Jun. 2019, doi: 10.1590/fst.02518.

[14] C. Nayak and V. K. Pathak, "Design and Development of Machine Fault Simulator (Mfs) for Fault Diagnosis," *Int. J. Recent Adv. Mech. Eng.*, vol. 4, no. 4, pp. 77–84, Nov. 2015, doi: 10.14810/ijmech.2015.4407.

[15] A. Alsubayhin, M. S. Ramzan, and B. Alzahrani, "Crime Prediction Model using Three Classification Techniques: Random Forest, Logistic Regression, and LightGBM," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 1, pp. 240–251, 2024, doi: 10.14569/IJACSA.2024.0150123.

[16] J. Sandhu, B. Mahapatra, S. Kulkarni, and A. Bhatt, "MetroPT Predictive Maintenance Using Logistic Regression and Random Forest with Isolation Forest Preprocessing," in *Lecture Notes in Networks and Systems*, P. M., D. K., and N. A., Eds., School of Data Science, Symbiosis Skills and Professional University, Kiwale, Maharashtra, Pune, 412101, India: Springer Science and Business Media Deutschland GmbH, 2024, pp. 503–513. doi: 10.1007/978-981-97-3180-0_33.

[17] M. H. Moharam, O. Hany, A. Hany, A. Mahmoud, M. Mohamed, and S. Saeed, "Anomaly detection using machine learning and adopted digital twin concepts in radio environments," *Sci. Rep.*, vol. 15, no. 1, p. 18352, May 2025, doi: 10.1038/s41598-025-02759-5.

[18] S. Kim, H. Seo, and E. C. Lee, "Advanced Anomaly Detection in Manufacturing Processes: Leveraging Feature Value Analysis for Normalizing Anomalous Data," *Electronics*, vol. 13, no. 7, p. 1384, Apr. 2024, doi: 10.3390/electronics13071384.

[19] C. El Morr, M. Jammal, H. Ali-Hassan, and W. El-Hallak, "Decision Trees," in *International Series in Operations Research and Management Science*, vol. 334, School of Health Policy and Management, York University, Toronto, ON, Canada: Springer, 2022, pp. 251–278. doi: 10.1007/978-3-031-16990-8_8.

[20] M. K. Gangwar, N. K. Sahu, R. K. Shukla, and C. Nayak, "Machine learning based progressive crack fault monitoring on spur gear using vibration analysis," *Int. J. Veh. Noise Vib.*, vol. 20, no. 1, pp. 89–106, Jan. 2024, doi: 10.1504/IJVNV.2024.138126.

[21] R. R. Saddala, D. Radha, and V. S. K. Devi, "Predictive Maintenance Analysis Using Multivariate Machine Performance Data for Industrial Downtime Reduction," in *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, Department of Computer Science and Engineering, Amrita school of computing, Amrita Vishwa Vidyapeetham, Bengaluru, India: IEEE, Feb. 2025, pp. 2283–2288. doi: 10.1109/IDCIOT64235.2025.10915129.

[22] S. Padmavathi, B. Nevetha, R. B. Sakthi, and K. M. A. Varshini, "Machine Learning for Reliable Industrial Operations," in *Cybersecurity and Data Science Innovations for Sustainable Development of HEICC*, Boca Raton, Thiagarajar College of Engineering, Tamilnadu, Madurai, India: CRC Press, 2024, pp. 316–330. doi: 10.1201/9781032711300-22.

[23] M. Gao, Y. Wei, Z. Li, B. Huang, C. Zheng,

and A. Mulati, "A Survey of Machine Learning Algorithms for Defective Steel Plates Classification," in *Lecture Notes in Electrical Engineering*, vol. 1252 LNEE, S. S. Y., Ed., College of Engineering and Computing, Florida International University, Miami, United States: Springer Science and Business Media Deutschland GmbH, 2024, pp. 467–476. doi: 10.1007/978-981-97-6934-6_55.

[24] M. Kopp, T. Pevný, and M. Holeňa, "Anomaly explanation with random forests," *Expert Syst. Appl.*, vol. 149, p. 113187, Jul. 2020, doi: 10.1016/j.eswa.2020.113187.

[25] Y. Cao and Y. Liang, "A Logistic Regression-based Early Warning System for Financial Risks under the Digital Transformation of Manufacturing Enterprises," in *2022 6th Annual International Conference on Data Science and Business Analytics (ICDSBA)*, IEEE, Oct. 2022, pp. 278–285. doi: 10.1109/ICDSBA57203.2022.00032.

[26] J. A. Martínez Pérez and P. S. Pérez Martín, "Regresión logística," *Med. Fam. Semer.*, vol. 50, no. 1, p. 102086, 2024, doi: 10.1016/j.semerg.2023.102086.

[27] M. Batra and R. Agrawal, "Comparative Analysis of Decision Tree Algorithms," B. K. Panigrahi, M. N. Hoda, V. Sharma, and S. Goel, Eds., Singapore: Springer Singapore, 2018, pp. 31–36. doi: 10.1007/978-981-10-6747-1_4.

[28] R. Buaton *et al.*, "Decision Tree Optimization in Data Mining with Support and Confidence," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, p. 012056, Aug. 2019, doi: 10.1088/1742-6596/1255/1/012056.

[29] Y. Manzali and M. Elfar, "Random Forest Pruning Techniques: A Recent Review," *Oper. Res. Forum*, vol. 4, no. 2, p. 43, May 2023, doi: 10.1007/s43069-023-00223-6.

[30] A. B. Shaik and S. Srinivasan, "A Brief Survey on Random Forest Ensembles in Classification Model," S. Bhattacharyya, A. E. Hassanien, D. Gupta, A. Khanna, and I. Pan, Eds., Singapore: Springer Singapore, 2019, pp. 253–260. doi: 10.1007/978-981-13-2354-6_27.

[31] K. Shukla, S. Nefti-Meziani, and S. Davis, "A heuristic approach on predictive maintenance techniques: Limitations and scope," *Adv. Mech. Eng.*, vol. 14, no. 6, pp. 1–14, Jun. 2022, doi: 10.1177/16878132221101009.

[32] A. Kumar, P. Sathujoda, and N. A. Bhalla, "Vibration signal analysis of a rotor-bearing system through wavelet transform and empirical mode decomposition," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1248, no. 1, p. 012027, Jul. 2022, doi: 10.1088/1757-899X/1248/1/012027.

[33] A. M. Umbrajkaar, A. Krishnamoorthy, and R. B. Dhumale, "Vibration Analysis of Shaft Misalignment Using Machine Learning Approach under Variable Load Conditions," *Shock Vib.*, vol. 2020, pp. 1–12, Jul. 2020, doi: 10.1155/2020/1650270.

[34] A. Jaramillo-Alcazar, J. Govea, and W. Villegas-Ch, "Anomaly Detection in a Smart Industrial Machinery Plant Using IoT and Machine Learning," *Sensors*, vol. 23, no. 19, p. 8286, Oct. 2023, doi: 10.3390/s23198286.

[35] M. Gamal, A. Donkol, A. Shaban, F. Costantino, G. Di Gravio, and R. Patriarca, "Anomalies Detection in Smart Manufacturing Using Machine Learning and Deep Learning Algorithms," *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, pp. 1611–1622, 2021, [Online]. Available: https://ieomsociety.org/proceedings/2021rome/573.pdf

[36] M. Huang, Z. Liu, and Y. Tao, "Mechanical fault diagnosis and prediction in IoT based on multi-source sensing data fusion," *Simul. Model. Pract. Theory*, vol. 102, no. September 2019, p. 101981, Jul. 2020, doi: 10.1016/j.simpat.2019.101981.

[37] D. Bagci Das and O. Das, "GABoT: A Lightweight Real-Time Adaptable Approach for Intelligent Fault Diagnosis of Rotating Machinery," *J. Vib. Eng. Technol.*, vol. 12, no. S1, pp. 679–697, Dec. 2024, doi: 10.1007/s42417-024-01440-x.

[38] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, "Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests," *IEEE Sens. J.*, vol. 17, no. 17, pp. 5581–5588, Sep. 2017, doi: 10.1109/JSEN.2017.2726011.

[39] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mech. Syst. Signal Process.*, vol. 76–77, pp. 283–293, Aug. 2016, doi: 10.1016/j.ymssp.2016.02.007.

[40] J. K. Chow, Z. Su, J. Wu, P. S. Tan, X. Mao, and Y. H. Wang, "Anomaly detection of defects on concrete structures with the convolutional autoencoder," *Adv. Eng. Informatics*, vol. 45, no. February, p. 101105, Aug. 2020, doi: 10.1016/j.aei.2020.101105.

[41] L. Scime and J. Beuth, "A multi-scale convolutional neural network for autonomous anomaly detection and classification in a laser powder bed fusion additive manufacturing process," *Addit. Manuf.*, vol. 24, pp. 273–286, Dec. 2018, doi: 10.1016/j.addma.2018.09.034.

[42] C. Cooper, J. Zhang, R. X. Gao, P. Wang, and I. Ragai, "Anomaly detection in milling tools using acoustic signals and generative adversarial networks," *Procedia Manuf.*, vol. 48, no. 2019, pp. 372–378, 2020, doi: 10.1016/j.promfg.2020.05.059.

[43] G. Mattera and L. Nele, "Machine learning approaches for real-time process anomaly detection in wire arc additive manufacturing," *Int. J. Adv. Manuf. Technol.*, vol. 137, no. 5–6, pp. 2863–2888, Mar. 2025, doi: 10.1007/s00170-025-15327-y.

[44] E. G. Sperandio Nascimento, J. S. Liang, I. S. Figueiredo, and L. L. N. Guarieiro, "T4pdm: A Deep Neural Network Based on the Transformer Architecture for Fault Diagnosis of Rotating Machinery," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4267690.

[45] Q. Meng and S. Zhu, "Anomaly detection for construction vibration signals using unsupervised deep learning and cloud computing," *Adv. Eng. Informatics*, vol. 55, p. 101907, Jan. 2023, doi: 10.1016/j.aei.2023.101907.

[46] A. Althubaiti, F. Elasha, and J. A. Teixeira, "Fault diagnosis and health management of bearings in rotating equipment based on vibration analysis – a review," *J. Vibroengineering*, vol. 24, no. 1, pp. 46–74, Feb. 2022, doi: 10.21595/jve.2021.22100.

[47] M. Tiboni, C. Remino, R. Bussola, and C. Amici, "A Review on Vibration-Based Condition Monitoring of Rotating Machinery," *Appl. Sci.*, vol. 12, no. 3, p. 972, Jan. 2022, doi: 10.3390/app12030972.

[48] M. Romanssini, P. C. C. de Aguirre, L. Compassi-Severo, and A. G. Girardi, "A Review on Vibration Monitoring Techniques for Predictive Maintenance of Rotating Machinery," *Eng*, vol. 4, no. 3, pp. 1797–1817, Jun. 2023, doi: 10.3390/eng4030102.

[49] Y. Cheng, M. Lin, J. Wu, H. Zhu, and X. Shao, "Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network," *Knowledge-Based Syst.*, vol. 216, p. 106796, Mar. 2021, doi: 10.1016/j.knosys.2021.106796.

[50] A. Rafati and H. R. Shaker, "Predictive maintenance of district heating networks: A comprehensive review of methods and challenges," *Therm. Sci. Eng. Prog.*, vol. 53, no. July, p. 102722, Aug. 2024, doi: 10.1016/j.tsep.2024.102722.

[51] B. M. Erland, A. K. Thorpe, and J. A. Gamon, "Recent Advances Toward Transparent Methane Emissions Monitoring: A Review," *Environ. Sci. Technol.*, vol. 56, no. 23, pp. 16567–16581, Dec. 2022, doi: 10.1021/acs.est.2c02136.

[52] J. Johny, S. Amos, and R. Prabhu, "Optical Fibre-Based Sensors for Oil and Gas Applications," *Sensors*, vol. 21, no. 18, p. 6047, Sep. 2021, doi: 10.3390/s21186047.

[53] M. Wang, X. Zhu, G. Zhou, K. Li, Q. Wu, and W. Fan, "Anomaly detection in multidimensional time series for water injection pump operations based on LSTMA-AE and mechanism constraints," *Sci. Rep.*, vol. 15, no. 1, p. 2020, Jan. 2025, doi: 10.1038/s41598-025-85436-x.

[54] B. Zhou *et al.*, "Smart table surface: A novel approach to pervasive dining monitoring," in *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, IEEE, Mar. 2015, pp. 155–162. doi: 10.1109/PERCOM.2015.7146522.

[55] H. Kim and I. Joe, "Enhancing Anomaly Detection in Maritime Operational IoT Time Series Data with Synthetic Outliers," *Electronics*, vol. 13, no. 19, p. 3912, Oct. 2024, doi: 10.3390/electronics13193912.

[56] Z. Hu, W. Chen, H. Wang, P. Tian, and D. Shen, "Integrated data-driven framework for anomaly detection and early warning in water distribution system," *J. Clean. Prod.*, vol. 373, p. 133977, Nov. 2022, doi: 10.1016/j.jclepro.2022.133977.

[57] J. Rabatel, S. Bringay, and P. Poncelet, "Anomaly detection in monitoring sensor data for preventive maintenance," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7003–7015, Jun. 2011, doi: 10.1016/j.eswa.2010.12.014.

[58] H. Farahani, R. Wagiran, and M. Hamidon, "Humidity Sensors Principle, Mechanism, and Fabrication Technologies: A Comprehensive Review," *Sensors*, vol. 14, no. 5, pp. 7881–7939, Apr. 2014, doi: 10.3390/s140507881.

[59] M. Pech, J. Vrchota, and J. Bednář, "Predictive Maintenance and Intelligent Sensors in Smart Factory: Review," *Sensors*, vol. 21, no. 4, p. 1470, Feb. 2021, doi: 10.3390/s21041470.

[60] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.

[61] C. F. Dormann *et al.*, "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance," *Ecography (Cop.).*, vol. 36, no. 1, pp. 27–46, Jan. 2013, doi: 10.1111/j.1600-0587.2012.07348.x.

[62] P. Turner, "Critical values for the Durbin-Watson test in large samples," *Appl. Econ. Lett.*, vol. 27, no. 18, pp. 1495–1499, Oct. 2020, doi: 10.1080/13504851.2019.1691711.

[63] L. Barreñada, P. Dhiman, D. Timmerman, A.-L. Boulesteix, and B. Van Calster, "Understanding random forests and overfitting: a visualization and simulation study," *Diagnostic Progn. Res.*, pp. 1–14, 2024, doi: 10.1186/s41512-024-00177-1.

[64] E. Halabaku and E. Bytyçi, "Overfitting in Machine Learning: A Comparative Analysis of Decision Trees and Random Forests," *Intell. Autom. Soft Comput.*, vol. 39, no. 6, pp. 987–1006, 2024, doi: 10.32604/iasc.2024.059429.

[65] A. Borucka and M. Grzelak, "Application of logistic regression for production machinery effciency evaluation," *Appl. Sci.*, vol. 9, no. 22, 2019, doi: 10.3390/app9224770.

[66] Q. Hu, X. S. Si, Q. H. Zhang, and A. S. Qin, "A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests," *Mech. Syst. Signal Process.*, vol. 139, p. 106609, 2020, doi: 10.1016/j.ymssp.2019.106609.

[67] M. Jiang, T. Cheng, K. Dong, S. Xu, and Y. Geng, "Fault diagnosis method of submersible screw pump based on random forest," *PLoS One*, vol. 15, no. 11, p. e0242458, Nov. 2020, doi: 10.1371/journal.pone.0242458.

[68] J. Ding, Q. Luo, L. Jia, and J. You, "Deep Forest-Based Fault Diagnosis Method for Chemical Process," *Math. Probl. Eng.*, vol. 2020, pp. 1–15, Jan. 2020, doi: 10.1155/2020/5281512.

[69] W. Wen, J. Qin, X. Xu, K. Mi, and M. Zhou, "A Model-Driven Approach to Extract Multi-Source Fault Features of a Screw Pump," *Processes*, vol. 12, no. 11, p. 2571, Nov. 2024, doi: 10.3390/pr12112571.

[70] R. B. Randall, *Vibration-based condition monitoring: industrial, aerospace, and automotive applications*. United States: Wiley, 2011. doi: 10.1002/9780470977668.

[71] N. Tambake, B. Deshmukh, S. Pardeshi, S. Salunkhe, R. Cep, and E. A. Nasr, "Fault diagnosis of a CNC hobbing cutter through machine learning using three axis vibration data," *Heliyon*, vol. 11, no. 2, p. e41637, Jan. 2025, doi: 10.1016/j.heliyon.2025.e41637.

[72] Z. Yuan, T. Zhou, J. Liu, C. Zhang, and Y. Liu, "Fault Diagnosis Approach for Rotating Machinery Based on Feature Importance Ranking and Selection," *Shock Vib.*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/8899188.

[73] O. Das, D. Bagci Das, and D. Birant, "Machine learning for fault analysis in rotating machinery: A comprehensive review," *Heliyon*, vol. 9, no. 6, p. e17584, Jun. 2023, doi: 10.1016/j.heliyon.2023.e17584.

[74] Y. Wei, Y. Li, M. Xu, and W. Huang, "A Review of Early Fault Diagnosis Approaches and Their Applications in Rotating Machinery," *Entropy*, vol. 21, no. 4, p. 409, Apr. 2019, doi: 10.3390/e21040409.

[75] G. Lin, M. Pritoni, Y. Chen, and J. Granderson, "Development and Implementation of Fault-Correction Algorithms in Fault Detection and Diagnostics Tools," *Energies*, vol. 13, no. 10, p. 2598, May 2020, doi: 10.3390/en13102598.

[76] M. Achouch *et al.*, "On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges," 2022. doi: 10.3390/app12168081.

[77] S. Zheng and Y. Zhou, "Digital Transformation and Competitive Advantage in Manufacturing: The Role of Business Model Innovation," 2025. doi: 10.3390/economies13070209.